What is Claimed:

1. A method for creating a task dependent unified language model for a selected application from a task independent corpus, the task dependent unified language model being for use in a language processing system and having embedded context-free grammar non-terminal tokens in a N-gram model, the method comprising:

obtaining a plurality of context-free grammars comprising non-terminal tokens representing semantic or syntactic concepts, each of the context-free grammars having words present in the task independent corpus to form the semantic or syntactic concepts;

parsing the task independent corpus with the plurality of context-free grammars to identify word occurrences of each of the semantic or syntactic concepts;

replacing each of the identified word occurrences with corresponding non-terminal tokens;

building a N-gram model having the non-terminal tokens; and

obtaining a second plurality of context-free grammars comprising at least some of the same non-terminals representing the same semantic or syntactic concepts, each of the context-free grammars of the second plurality being more appropriate for use in the selected application.

2. The method of claim 1 and further comprising:

storing the N-gram model having the non-terminal tokens and the second plurality of context-free grammars having non-terminal tokens representing task dependent semantic or syntactic concepts on a computer readable medium.

3. A method for creating a task dependent unified language model for a selected application from a task independent corpus, the task dependent unified language model being for use in a language processing system and having embedded context-free grammar non-terminal tokens in a N-gram model, the method comprising:

obtaining a plurality of context-free grammars comprising a set of context-free grammars having non-terminal tokens representing task dependent semantic or syntactic concepts and at least one context-free grammar having a non-terminal token for a phrase that can be mistaken for one of the desired task dependent semantic or syntactic concepts;

parsing the task independent corpus with the plurality of context-free grammars to identify word occurrences for each of the semantic or syntactic concepts and phrases;

replacing each of the identified word occurrences with corresponding non-terminal tokens; and

building a N-gram model having the non-terminal tokens.

4. The method of claim 3 wherein replacing each of the identified word occurrences includes excluding the non-

terminals added for the prevention of mistakes during parsing.

5. The method of claim 3 and further comprising:

storing the N-gram model having the non-terminal tokens and the set of context-free grammars having non-terminal tokens representing task dependent semantic or syntactic concepts on a computer readable medium.

6. The method of claim 3 wherein building the N-gram model includes eliminating at least some of the associated text from the task independent corpus for non-terminal tokens that can be mistaken for one of the desired task dependent semantic or syntactic concepts.

7. A method for creating a language model for a selected application from a task independent corpus, the language model being for use in a language processing system, the method comprising:

obtaining a plurality of context-free grammars comprising non-terminal tokens representing semantic or syntactic concepts of the selected application;

generating word phrases from the plurality of context-free grammars;

formulating an information retrieval query from at least one of the word phrases;

querying the task independent corpus based on the query formulated;

identifying associated text in the task independent corpus based on the query; and

building a language model using the identified text.

8. The method of claim 7 wherein building a language model comprises building a N-gram language model.

9. The method of claim 8 and further comprising:

parsing the identified text of the task independent corpus with the plurality of context-free grammars to identify word occurrences for each of the semantic or syntactic concepts;

replacing each of the identified word occurrences with corresponding non-terminal tokens; and

wherein building the N-gram language model comprises building a N-gram model having the non-terminal tokens.

10. The method of claim 8 and further comprising:

building a second N-gram language model from the word phrases generated from the plurality of context-free grammars; and

combining the first-mentioned N-gram language model and the second N-gram language model to form a third N-gram language model.

11. The method of claim 10 and further comprising:

parsing the identified text of the task independent corpus with the plurality of context-free grammars to identify word occurrences for each of the semantic or syntactic concepts;

replacing each of the identified word occurrences with corresponding non-terminal tokens; and

wherein the word phrases include non-terminal tokens and wherein building the first-mentioned N-gram language model comprises building a N-gram model having the non-terminal tokens.

12. The method of claim 9 and further comprising:

storing the N-gram model having the non-terminal tokens and the plurality of context-free grammars having non-terminal tokens representing task dependent semantic or syntactic concepts on a computer readable medium.

13. The method of claim 7 and further comprising:

storing the identified text of the task independent corpus separate from the task independent corpus.

14. A method for creating a language model for a selected application from a task independent corpus, the language model being for use in a language processing system, the method comprising:

obtaining a plurality of context-free grammars comprising non-terminal tokens representing semantic or syntactic concepts of the selected application;

generating word phrases from the plurality of context-free grammars;

building a first N-gram language model from the word phrases;

formulating an information retrieval query from at least one of the word phrases;

querying the task independent corpus based on the query formulated;

identifying associated text in the task independent corpus based on the query; and

building a second N-gram language model from the identified text; and

combining the first N-gram language model and the second N-gram language model to form a third N-gram language model.

15. The method of claim 14 wherein building the second N-gram language model includes using only the identified text.

16. The method of claim 15 and further comprising:

storing the identified text of the task independent corpus separate from the task independent corpus.

17. The method of claim 16 and further comprising:

parsing the identified text of the task independent corpus with the plurality of context-free grammars to identify word occurrences for each of the semantic or syntactic concepts;

replacing each of the identified word occurrences with corresponding non-terminal tokens; and

wherein the word phrases include non-terminal tokens and wherein building the first-mentioned N-gram language model comprises building a N-gram model having the non-terminal tokens.

18. The method of claim 14 and further comprising:

parsing the task independent corpus with the plurality of context-free grammars to identify word

occurrences for each of the semantic or syntactic concepts;

replacing each of the identified word occurrences with corresponding non-terminal tokens; and

wherein the word phrases include non-terminal tokens and wherein building the first-mentioned N-gram language model comprises building a N-gram model having the non-terminal tokens.

19. A method for creating a unified language model for a selected application from a corpus, the method comprising:

obtaining a plurality of context-free grammars comprising non-terminal tokens representing semantic or syntactic concepts of the selected application;

building a word language model from the corpus; and

assigning probabilities to words of at least some of the context-free grammars as a function of corresponding probabilities obtained for the same words from the word language model wherein assigning probabilities includes normalizing the probabilities of the words from the language model in each of the context-free grammars as a function of the words allowed by the corresponding context-free grammar.

20. The method of claim 19 wherein the word language model comprises an N-gram language model.

21. The method of claim 19 wherein the corpus comprises a task independent corpus.

22. The method of claim 21 and further comprising:

generating word phrases from the plurality of context-free grammars;

formulating an information retrieval query from at least one of the word phrases;

querying the task independent corpus based on the query formulated;

identifying associated text in the task independent corpus based on the query; and

wherein building a N-gram language model includes using the identified text.

23. A computer readable medium including instructions readable by a computer which, when implemented execute a method to build a task dependent unified language model for a language processing system, the method comprising:

accessing a plurality of context-free grammars comprising non-terminal tokens representing semantic or syntactic concepts, each of the context-free grammars having words present in a task independent corpus to form the semantic or syntactic concepts;

parsing the task independent corpus with the plurality of context-free grammars to identify word occurrences of each of the semantic or syntactic concepts;

replacing each of the identified word occurrences with corresponding non-terminal tokens;

building a N-gram model having the non-terminal tokens; and

storing the N-gram model and a second plurality of context-free grammars comprising at least some of

the same non-terminals representing the same semantic or syntactic concepts, each of the context-free grammars of the second plurality being more appropriate for use in the selected application.

24. A computer readable medium including instructions readable by a computer which, when implemented execute a method to build a task dependent unified language model for a language processing system, the method comprising:

accessing a plurality of context-free grammars comprising a set of context-free grammars having non-terminal tokens representing task dependent semantic or syntactic concepts and at least one context-free grammar having a non-terminal token for a phrase that can be mistaken for one of the desired task dependent semantic or syntactic concepts;

parsing a task independent corpus with the plurality of context-free grammars to identify word occurrences for each of the semantic or syntactic concepts and phrases;

replacing each of the identified word occurrences with corresponding non-terminal tokens; and

building a N-gram model having the non-terminal tokens.

25. The computer readable medium of claim 24 wherein replacing each of the identified word occurrences with corresponding non-terminal tokens includes excluding the non-terminals added for the prevention of mistakes during parsing.

26. The computer readable medium of claim 24 having instructions further comprising:

storing the N-gram model having the non-terminal tokens and the set of context-free grammars having non-terminal tokens representing task dependent semantic or syntactic concepts on a computer readable medium.

27. The computer readable medium of claim 24 wherein building the N-gram model includes eliminating at least some of the associated text from the task independent corpus for non-terminal tokens that can be mistaken for one of the desired task dependent semantic or syntactic concepts.

28. A computer readable medium including instructions readable by a computer which, when implemented execute a method to build language model for a language processing system, the method comprising:

accessing a plurality of context-free grammars comprising non-terminal tokens representing semantic or syntactic concepts of the selected application;

generating word phrases from the plurality of context-free grammars;

formulating an information retrieval query from at least one of the word phrases;

querying a task independent corpus based on the query formulated;

identifying associated text in the task independent corpus based on the query; and

building a language model using the identified text.

29. The computer readable medium of claim 28 wherein building a language model comprises building a N-gram language model.

30. The computer readable medium of claim 29 and having instructions further comprising:

parsing the identified text of the task independent corpus with the plurality of context-free grammars to identify word occurrences for each of the semantic or syntactic concepts;

replacing each of the identified word occurrences with corresponding non-terminal tokens; and

wherein building the N-gram language model comprises building a N-gram model having the non-terminal tokens.

31. The computer readable medium of claim 29 and having instructions further comprising:

building a second N-gram language model from the word phrases from the plurality of context-free grammars; and

combining the first-mentioned N-gram language model and the second N-gram language model to form a third N-gram language model.

32. The computer readable medium of claim 31 and having instructions further comprising:

parsing the identified text of the task independent corpus with the plurality of context-free grammars to identify word occurrences for each of the semantic or syntactic concepts;

replacing each of the identified word occurrences with corresponding non-terminal tokens; and

wherein the word phrases include non-terminal tokens and wherein building the first-mentioned N-gram language model comprises building a N-gram model having the non-terminal tokens.

33. The computer readable medium of claim 30 and having instructions further comprising:

storing the N-gram model having the non-terminal tokens and the plurality of context-free grammars having non-terminal tokens representing task dependent semantic or syntactic concepts on a computer readable medium.

34. The computer readable medium of claim 28 and having instructions further comprising:

storing the identified text of the task independent corpus separate from the task independent corpus.

35. A computer readable medium including instructions readable by a computer which, when implemented execute a method to build language model for a language processing system, the method comprising:

accessing a plurality of context-free grammars comprising non-terminal tokens representing semantic or syntactic concepts of the selected application;

generating word phrases from the plurality of context-free grammars;

building a first N-gram language model from the word phrases;

formulating an information retrieval query from at least one of the word phrases;

querying a task independent corpus based on the query formulated;

identifying associated text in the task independent corpus based on the query;

building a second N-gram language model from the identified text; and

combining the first N-gram language model and the second N-gram language model to form a third N-gram language model.

36. The computer readable medium of claim 35 wherein building the second N-gram language model includes using only the identified text.

37. The computer readable medium of claim 36 and having instructions further comprising:

storing the identified text of the task independent corpus separate from the task independent corpus.

38. The computer readable medium of claim 37 and having instructions further comprising:

parsing the identified text of the task independent corpus with the plurality of context-free grammars to identify word occurrences for each of the semantic or syntactic concepts;

replacing each of the identified word occurrences with corresponding non-terminal tokens; and

wherein the word phrases include non-terminal tokens and wherein building the first-mentioned N-gram

language model comprises building a N-gram model having the non-terminal tokens.

39. The computer readable medium of claim 35 and having instructions further comprising:

parsing the task independent corpus with the plurality of context-free grammars to identify word occurrences for each of the semantic or syntactic concepts;

replacing each of the identified word occurrences with corresponding non-terminal tokens; and

wherein the word phrases include non-terminal tokens and wherein building the first-mentioned N-gram language model comprises building a N-gram model having the non-terminal tokens.

40. A computer readable medium including instructions readable by a computer which, when implemented execute a method to build a unified language model for a selected application, the method comprising:

accessing a plurality of context-free grammars comprising non-terminal tokens representing semantic or syntactic concepts of the selected application;

building a word language model from a corpus; and

assigning probabilities to words of at least some of the context-free grammars as a function of corresponding probabilities obtained for the same terminals from the word language model wherein assigning probabilities includes normalizing the probabilities of the words from the word language model in each of the context-free grammars as a

function of the words allowed by the corresponding context-free grammar.

41. The computer readable medium of claim 40 wherein the word language model comprises an N-gram language model.

42. The computer readable medium of claim 40 wherein the corpus comprises a task independent corpus.

43. The computer readable medium of claim 42 and having instructions further comprising:

generating word phrases from the plurality of context-free grammars;

formulating an information retrieval query from at least one of the word phrases;

querying the task independent corpus based on the query formulated;

identifying associated text in the task independent corpus based on the query; and

wherein building a N-gram language model includes using the identified text.